X EXECUTIVE SUMMARY

X.1 INTRODUCTION

X.1.1 Background on this Project

During the electricity crises of the last two years, a number of states and utilities within these states have developed programs to encourage customers to reduce their peak loads on short notice (under 2 to 24 hours) in exchange for some form of compensation. Such demand response (DR) programs depend on a credible operational procedure for determining the magnitude of load reductions for each customer during each load reduction period.

The use of inconsistent methods for calculating baselines and corresponding load reductions has caused both confusion and dissatisfaction among participating customers. The lack of a standard measurement procedure may be reducing the number of customers willing to participate in DR programs, particularly in smaller- and medium-sized commercial customers in California.

X.1.2 Objectives

The objective of this work is to develop a standardized measurement and verification (M&V) protocol for use by building engineers, facility operators, or outside M&V experts to "measure" the load drops achieved at a premises. Completion of this protocol is aimed at increasing participation in DR programs from small- and medium-sized customers by reducing the barriers related to inconsistency and confusion about baseline methods.

X.1.3 Project Steps

Steps in the project include:

- Review of existing methods
- Testing of alternative methods on data sets from various locations and customer types
- Draft report on findings and recommendations, circulated for review and presented for discussion at a public workshop
- Final report
- Submission of the final recommendations to the International Performance Measurement and Verification Protocol (IPMVP) organization for adoption as part of the IPMVP.

This is the final report on findings and recommendations. Included are the review of existing methods and the results of tests on alternative methods.

X.1.4 The Role of the IPMVP

The IPMVP organization has participated in the development of this report and recommendations. The organization is responsible for the continued development and dissemination of standardized verification methods. It is hoped that the involvement of the IPMVP at various stages of review and the anticipated adoption of the DR protocol as an IPMVP document will represent a broad base of support for the framework developed.

There are direct parallels in the current demand response area to what was occurring in the world of M&V for energy-efficiency performance contracting eight years ago. The core concept of the IPMVP document is that parties involved in contracts to reduce energy use should have a common language with which to structure and manage the settlement of those contracts. The IPMVP was designed to allow parties flexibility in designing M&V procedures that make sense for each contact.

As is true for the energy-efficiency IPMVP, the intent of this report is not to provide a prescriptive set of steps and rules. Rather, the goal is to establish a clear vocabulary, and to offer guidelines on good practice and the pros and cons of alternative method specifications. Toward the goal of developing consistent terminology, this document develops a taxonomy of different methods and attempts to provide clear definitions. We anticipate that a discussion of definitions and distinctions will be an important part of the refinement of this document.

X.1.5 The Role of Other Contributors

This work would not have been possible without the contributions of several other organizations and individuals.

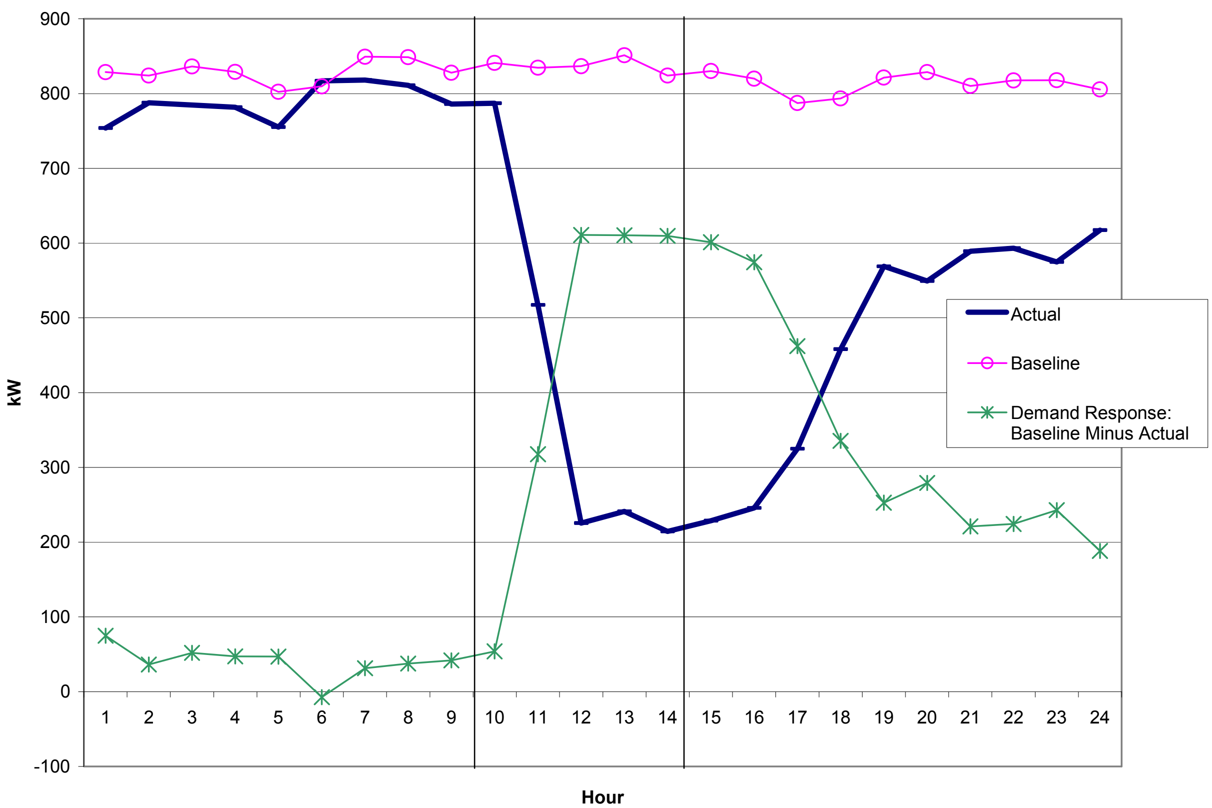
- *Method donors* have shared details of methods they have developed and applied for quantifying demand reductions.
- *Data donors* have provided interval load data from curtailed and noncurtailed customers for use in testing methods.
- *Reviewers* who reviewed the major project deliverables.

X.2 Review of Existing Methods

X.2.1 *The Demand Response Baseline*

The focus of this study was on calculations of demand response from whole-premise interval load data. Demand response is calculated as the difference between the *baseline* and the actual metered load in each interval (Figure X-1). The baseline is the estimate of what the load would have been in each interval in the absence of the curtailment. Thus, the key question for the demand response calculation is how the baseline is determined.

Figure X-1
Example of Demand Response Calculation from Baseline and Actual Loads



X.2.2 Desirable Features in a Demand Response Baseline Calculation Method

Most of the goals developers described for the baseline were similar. They were

- to reflect load that would have been used absent the program,
- ease of use for program participants,

- ease of use for program administrators, and
- deterrence of gaming.

Given the multiple objectives, all customer baseline developers understood that the baseline methodology they chose was a compromise. Criteria that are balanced in developing a baseline include

- Simplicity
- Ease of use
- Ease of understanding
- Verifiability
- Accuracy
- Lack of bias (i.e., no systematic tendency to over- or under-state reductions)
- Ability to handle weather-sensitive accounts fairly
- Minimization of gaming
- Ability to be known prior to customer's commitment to a particular curtailment amount and event
- Costs for participant and operator to implement
- Consistency with other ISOs.

X.2.3 Components of Whole-Premise Baseline Calculation Methods

Baseline calculation methods based on whole-premise interval-metering data can be described in terms of three fundamental components:

- **Data selection criteria** determine what days and time periods of data will be used in the baseline calculation.
- The **estimation method** is a calculation procedure that determines the provisional baseline load at each interval for the curtailment day, using the data selected by the data selection criteria.
- The **adjustment method** shifts or scales the provisional baseline to align it with known conditions of the curtailment day.

Data Selection Criteria

Common starting points for data selection include

- Use of the last 10 to 20 uncurtailed business days
- Use of a subset of the last 10 or 11 business days that had the highest load

- Use of a full season of data.

Selection criteria include varying procedures for excluding days from the starting point and replacing excluded days, sometimes in an iterative process.

Estimation Method

Most estimation methods can be characterized as either an average or some form of weather-based regression model.

Adjustment Methods

Common adjustment methods (used to adjust interval meter on the day of the curtailment) include:

- Unadjusted
- Additive
- Scalar
- Weather-based additive or scalar.

X.2.4 Characterization of Existing Methods

Existing baseline estimation methods are summarized in Figure X-1 in terms of the components discussed above.

Figure X-2
Summary of Existing Baseline Estimation Methods

	Load Type Differences Addressed*	How Weather Sensitivity is Addressed**	Time Interval	Data Selection			Estimation Method	Adjustment Hours, If Adjusted
				Initial Timeframe	Final Selection	Excluded Days (other than weekends and previous program Control days)		
PJM-Day Ahead 2002	WS/NWS, Self Gen, Cust. Spec.	Top 5 of 10, Optional adjustment to control-day weather	Hourly	10 days, n-2 to n-11	High 5 of 10	Low Output Days.	Interval Average	h-1,h-2.
PJM Emergency	None	None	Hourly	Hour before	Same	None		
ISO-NE 2001-2002	WS/NWS, Self Gen,	Adjustment to control-day load	Hourly	10 days, n-1 to n-10	Same	Extreme Output Days.	Interval Average	h-1,h-2.
NYISO-DADRP 2001	Self Gen	Top 5 of 10	Hourly	10 days, n-2 to n-11	High 5 of 10	Low Output Days.	Interval Average	
NYISO-DADRP 2002	WS/NWS, Self Gen	Top 5 of 10, Optional adjustment to control-day load	Hourly	10 days, n-2 to n-11	High 5 of 10	Low Output Days.	Interval Average	h-3, h-4.
ERCOT-BUL 2002	WS/NWS	Optional adjustment to control-day load	15 minute	10 days, n-1 to n-10	Middle 8	None	Interval Average	h-1, h-2.
CAISO 2001#1	None	None	Hourly	10 days	None	None	Interval Average	
CAISO 2001#2	None	None	Hourly	11 days	None	None	Interval Average	
XENERGY	WS/NWS	Regression-based estimate, Adjustment to control-day load	Hourly	Variable	None	None	Regression-based	h-1, h-2.
LBNL/Kinney	WS/NWS	Regression-based estimate	Hourly	10 days, n-1 to n-10	None	None	Regression-based	
Nexant	WS/NWS	Adjustment to control-day load	15 minute	10 days	None	None	Interval Average	h-1
Utility A	WS/NWS	Adjustment to control-day load	Hourly	Previous Month	None	All Days that do not fit the match-day criteria.	Interval Average	one hour, 8am - 11am
Utility B	None	None	Hourly	5 days	None	Customer-specificed anomalous loads	Interval Average	
Utility C	WS/NWS	Regression-based estimate	Hourly	Undefined minimum data	None	None	Regression-based	
Utility D	WS/NWS	Regression-based estimate	Hourly	Weekdays, June through September	None	None	Regression-based	5am - 10am
Utility E	None	Match based data selection	15 minute	Undefined	10 Days with min. SSE compared to day n-1	None	Interval Average	All match-day hours.
Utility F	WS/NWS, Cust. Spec.	Adjustment to control-day load	Hourly	2-3 previous years	None	Anomalous loads	Interval Average	h-1, h-2.
CMTA Proposed OBMC	WS/NWS	Adjustment to control-day load	Hourly	10 days, n-1 to n-10	None	None	Interval Average	h-1 through h-4

* WS/NWS: Different methods for weather-sensitive and nonweather-sensitive loads
Self Gen: Different methods for onsite generation

** Top 5 of 10: Select 5 days with highest average load during the hours curtailed on the curtailment day

X.3 FINDINGS FROM METHOD TESTS

Several combinations of data selection criteria, estimation method, and adjustment were tested on interval load data from curtailed and uncurtailed customer accounts across the country. A total of 646 accounts were included in the tests. For accounts that were not curtailed, baseline estimates were compared with actual load for each hour of an actual or simulated curtailment period. For accounts that were curtailed, each candidate baseline estimate was compared with the estimate produced by the "best" method.

Performance of each method was assessed in terms of both bias and overall error magnitude. Bias is the systematic tendency to over- or under-state the baseline and corresponding demand reduction. Variability is how wide the swings are around the typical or expected value. Overall error magnitude reflects both bias and variability.

Key findings are indicated below. These findings indicate the effects of various method features on bias and variability as measured in this study for the accounts and specific methods tested. These results offer general guidelines, but the performance of a particular method in a particular situation may be different.

X.3.1 Adjustments

- Additive adjustment to the load data from two hours before curtailment can often reduce the bias and variability of almost all methods, including weather models, for weather-sensitive or non-weather-sensitive, high or low variability accounts. Other types of adjustments can improve the performance of averages, but generally with higher bias and variability.
- With this additive adjustment, simple averaging methods in most cases perform essentially as well as complex weather models, even for weather-sensitive accounts.
- Without adjustment, most averages tend to understate the load impacts of a curtailment.
- Additive adjustment to the last 2 hours before a curtailment can be problematic for several reasons:
  - ① It opens the possibility of gaming by deliberately increasing load just before the curtailment period to boost the baseline.
  - ① Legitimate pre-cooling in response to a curtailment notice or expectation will also erroneously increase the baseline.
  - ① Conversely, an operation that achieves its curtailment target promptly upon notification and before the beginning of the required curtailment period will have a severely understated baseline.

X.3.2 Data Selection

- Bias and variability of weather models tends to be reduced by the use of longer input data series, but not dramatically.
- The decreased variability with longer input series is more noticeable for conditional weather models applied to non-weather-sensitive accounts, particularly high-variability accounts.
- The different average methods performed similarly in terms of bias and variability, except for those that select a subset of days based on high load.
- For summer loads, the High 5 of 10 average generally reduces the otherwise negative bias. For summer loads using additive adjustment, the High 5 of 10 days gave the lowest bias measure of any of the averages, for both weather-sensitive and non-weather-sensitive accounts, and comparable variability. The High 10 of 11 average method gave some bias reduction, but not as much.
- For nonsummer loads, however, the High 5 of 10 average method inflates an already positive bias. The other averages perform better and roughly comparably to each other, in terms of both bias and variability, for both weather-sensitive and non-weather-sensitive accounts. The High 10 of 11 is somewhat better than the others in terms of the bias and variability measured in this study.

X.3.3 Weather Modeling

- For summer weather-sensitive accounts, weather models tend to perform somewhat better than averages, but the difference is not dramatic.
- For summer non-weather-sensitive accounts, use of a "conditional" weather model does not increase bias or variability. The conditional weather model automatically deletes weather terms if the statistical diagnostics based on the load data indicate these terms are inappropriate for a particular account. Use of such diagnostics protects against retaining terms in the model that are not well determined and are likely not to be meaningful. Thus, if weather models are used, a single methodology can be applied to both weather-sensitive and non-weather-sensitive accounts.
- For nonsummer loads, weather models do not perform better than averages.

X.4 PROS AND CONS OF ALTERNATIVE APPROACHES

Advantages and disadvantages of key method features in terms of the criteria indicated in Section X.2.2 are summarized in the table below. This table is based on both qualitative considerations and the results of the performance tests.

Table X-1
Advantages and Disadvantages of Key Baseline Method Features Based on Qualitative Considerations and Test Results

Baseline Method	Variant	Pros	Cons
Average	Any	Simple, easy to use and understand, low cost	Tends to understate baseline for weather-sensitive loads, especially if unadjusted
	High 5 of last 10 days	Partial adjustment for weather-sensitive loads	Still tends to understate baseline for weather-sensitive loads
			Can allow windfall load reduction credit on cool days
Regression	Any	Provides baseline corresponding to particular weather conditions of curtailment day	More complex, harder to understand, higher cost
			If observations don't include conditions as extreme as the curtailment day, model estimate may be inaccurate
			If account isn't weather-sensitive, may be less accurate than simpler methods
	Full Season	Adequate data and range of variation to yield accurate coefficients	Operating conditions from the period data are taken from may be different from curtailment day
	Recent 10 days	Operating conditions more likely to be similar to curtailment day	Model based on limited data may be inaccurate
	Lag temperature/ degree-day	Tends to reduce bias for weather-sensitive accounts	Tends to increase variability of baseline estimate.
	Conditional	Allows same general form and procedure to be used for weather-sensitive and non-weather-sensitive accounts, without pre-screening. Doesn't add much error for non-weather-sensitive accounts.	More complex. May give less consistent results across events for an account, if weather terms are sometimes retained and sometimes not.
Adjustment to precurtailment hours	Any	Simple, easy to use and understand, low cost	May be potential for gaming behavior during day-of-curtailment adjustment period
		Adjusts to weather and operating conditions of curtailment day	Appropriate pre-curtailment increase in load (e.g., pre-cooling) will result in overstated baseline
		Limits potential for collecting windfall credits for planned shut-downs	Pre-curtailment decrease in load in response to curtailment request (e.g., long ramp-down, canceling a shift) will result in understated baseline
	Additive	May adjust well for load change that is constant throughout day (e.g., industrial processes)	May not be appropriate if load changes during curtailment period (ratio adjustment may be better suited)
	Scalar	May adjust well for load change that is function of exogenous factor throughout day (e.g., higher levels of occupancy)	May not be appropriate if the day-to-day load variation is constant over the day (additive adjustment may be better suited)
	to last 2 hours before curtailment period	If load in these hours is unaffected by anticipated or initiated curtailment, provides best accuracy	If substantial curtailment is initiated in these hours, severely understates baselines
	to 3rd and 4th hour before curtailment period	Less potential for understated baseline due to pre-curtailment-period demand response	More variability than adjustment to last 2 hours
Weather-Based Adjustment	Any	Explicitly takes into account weather conditions	Adjustment may not be known to customer until after curtailment period (i.e., until after weather conditions are known for the day)
		No opportunity for gaming as with adjustment to precurtailment hours	If no observations are available for extreme conditions, estimates used for adjustment may be outside range of model
			Will badly predict load reductions if the buildings are dominated by internal loads
			Less accurate than alternative adjustments or weather model for both weather-sensitive and non-weather-sensitive accounts

X.5 RECOMMENDATIONS

In developing our recommendations, we did not attempt to score each method or feature with respect to each of the desirable features indicated above, nor assign explicit weights to the criteria. In general, our approach is

- allow for options that recognize different circumstances
- favor simplicity if the potential accuracy gains of greater complexity appear to be slight
- indicate alternatives and trade-offs with respect to the criteria.

X.5.1 Proposed Approaches by Account Type

Offering Options

A general recommendation is that baseline calculation protocols should provide for alternatives based on customer load characteristics and operating practices. One way to simplify the provision of options is to establish a default method and allow certain deviations.

The basis for the selection of a method should be not just the customer's business type, but also the load patterns evident in the data as well as the customer's description of operating practices. Thus, for example, a customer who indicates a desire to be able to cancel a shift in advance of the control period should have access to a baseline calculation method that is not distorted by this practice.

At the same time, the program operator should have some discretion to bar customers from using an approach that they appear to have manipulated in the past. Thus, if there is evidence that a particular customer tends to inflate the baseline load after notification, beyond what would reasonably be expected for pre-cooling, that customer might not be able to use a method that includes adjustment to the 2 pre-curtailment hours.

A Practical Default Baseline Calculation Method

A method that generally works well for a range of account types is the simple average of the last 10 days, with additive adjustment to the load shape 2 hours prior to the curtailment period. This method can be recommended for both weather-sensitive and non-weather-sensitive accounts, with both low and high variability, for summer and nonsummer curtailments.

This method is not recommended for accounts that tend to curtail in advance of the required period in response to a curtailment notice. It is also not recommended for situations where the potential for gaming is a strong concern, whether across the program or for particular customers.

Alternatives for Summer Weather-Sensitive Accounts

For summer programs, practical alternatives for weather-sensitive accounts include the following:

- Unadjusted weather models. Longer input time periods to estimate the baseline load are preferable, particularly for high-variability loads.
- The High 5 of 10 day average with Temperature-Humidity Index (THI) adjustment.

Simpler methods with less desirable but potentially acceptable performance include:

- Unadjusted averages, particularly the High 5 of 10.
- Averages or weather models adjusted to the third and fourth hour before curtailment.

Alternatives for Summer Non-Weather-Sensitive Accounts

For non-weather-sensitive summer loads, the unadjusted High 10 of 11 average performs nearly as well as the recommended default, particularly for low-variability loads. Next best is the simple average of the last 10 days with additive adjustment to the third and fourth hours before curtailment.

For low-variability loads, unadjusted weather models, with weather terms retained only if indicated by the data, actually perform slightly better than the recommended default. However, unlike the case for weather-sensitive accounts, these models perform better if based on shorter periods of data. For high-variability loads, unadjusted weather models tend to be worse than the unadjusted High 10 of 11 average.

Alternatives for Nonsummer Accounts

For nonsummer loads, modeling is more challenging and there are fewer alternatives. For weather-sensitive accounts, the High 5 of 10 day average with THI adjustment can be used. For low-variability loads, the unadjusted High 5 of 10 day appears to perform slightly better, but for high-variability loads it is worse.

For non-weather-sensitive nonsummer loads, the unadjusted High 10 of 11 appears to be the best alternative. Any of the averages with additive adjustment to the third and fourth hour before curtailment do not perform as well.

Summary of Recommended Methods and Alternatives

The recommended methods and alternatives for different account types are summarized in the table below.

Table X-2
Recommended Methods and Alternatives

Season	Weather Sensitivity	Variability	Recommended Default Estimation	Data Selection	Adjustment	Recommended Alternatives Estimation	Data Selection	Adjustment
Summer	Weather-Sensitive	Low	Average	last 10	add 1-2	weather models	any	none
						Average	High 5	THI
Summer	Weather-Sensitive	High	Average	last 10	add 1-2	weather models	longer is better	none
						Average	High 5 of 10	THI
Summer	Non-Weather-Sensitive	Low	Average	last 10	add 1-2	weather models	shorter is better	none
						Average	High 10 of 11	none
						Average	last 10	add 3-4
Summer	Non-Weather-Sensitive	High	Average	last 10	add 1-2	Average	High 10 of 11	none
						Average	last 10	add 3-4
Nonsummer	Weather-Sensitive	Low	Average	last 10	add 1-2	Average	High 5 of 10	none
Nonsummer	Weather-Sensitive	High	Average	last 10	add 1-2	Average	High 5 of 10	THI
Nonsummer	Weather-Sensitive	Low	Average	last 10	add 1-2	Average	High 10 of 11	none
						Average	last 10	add 3-4
Nonsummer	Weather-Sensitive	High	Average	last 10	add 1-2	Average	High 10 of 11	none
						Average	last 10	add 3-4

X.6 AREAS FOR FUTURE DEVELOPMENT

Reviewers of a draft of this report offered a number of valuable suggestions. Many of these suggestions have been incorporated in the final version. Others, while of considerable merit, were beyond the scope of what could be accomplished in this study.

Most of the suggestions that could not be addressed were in the following categories:

1. Test additional methods
2. Examine results by finer categories of customer type
3. Provide more explicit rules and clear-cut bases for choosing among alternatives, including methods for identifying gaming.

These issues will be explored in further work by the IPMVP Technical Committee. The committee will be building on this study to develop a DR baseline protocol that can be adopted as part of the IPMVP document. The Protocol itself will include the establishment of consistent terminology, guidance on appropriate methods for different situations, and rationale for that guidance. The technical analysis that forms a major portion of this report will not itself be part of the Protocol, but will be referenced as part of the rationale for the guidance.

As part of the further work by the IPMVP, some additional methods may be tested. Candidates for testing include:

- ASHRAE load forecasting models
- variable degree-day models
- use of a single hour or two hours prior to curtailment as a flat baseline.

The additional testing may also examine results by finer segments. Customer type is not known for most of the data sets examined in this study. However, customers can be classified by size, and possibly by other patterns in the load data.

An IPMVP Protocol by its nature will offer options and guidance rather than being prescriptive. However, this Protocol can serve as the basis for establishing specific rules and procedures within a jurisdiction, and provide a common language for understanding these procedures.